



Projeto e Implementação de Plataforma de Dados com Snowflake

ELT (Extração, Carga e Transformação) com Airbyte



Projeto e Implementação de Plataforma de Dados com Snowflake

O processo ELT (Extração, Carga e Transformação) é uma abordagem moderna para o gerenciamento de dados, onde os dados são extraídos de suas fontes, carregados em um destino central (geralmente um Data Warehouse ou Data Lake), e então transformados diretamente nesse destino. O Airbyte é uma ferramenta essencial para as etapas de extração e carga no contexto ELT.

Etapas do ELT com Airbyte

1. Extração (Extract)

Nesta etapa, os dados são retirados de suas fontes originais, que podem incluir bancos de dados, APIs, arquivos ou ferramentas SaaS (como Salesforce ou Google Analytics). O Airbyte se destaca por oferecer:

- Conectores prontos: Mais de 200 conectores que facilitam a integração com várias fontes de dados.
- Personalização: Permite criar conectores personalizados usando SDKs.
- Confiabilidade: Suporte para extração incremental, reduzindo a carga e o volume de dados processados.

2. Carga (Load)

Os dados extraídos são carregados no destino escolhido, como Snowflake, BigQuery, Redshift ou outros. O Airbyte executa essa etapa de forma eficiente:

- Formato bruto: Os dados são carregados no formato original, sem transformações, o que preserva a integridade dos dados.
- Escalabilidade: Suporta grandes volumes de dados, adequando-se a necessidades de empresas de diferentes portes.
- Monitoramento: Fornece logs detalhados e alertas para rastrear falhas no carregamento.

3. Transformação (Transform)

Embora o Airbyte seja focado em extração e carga, o processo de transformação ocorre posteriormente no destino. Ferramentas como dbt (Data Build Tool) são frequentemente utilizadas nesta etapa para:

- Modelar os dados conforme as necessidades do negócio.
- Aplicar regras de limpeza, padronização e agregação.
- Otimizar os dados para análise, criando tabelas prontas para relatórios e dashboards.

Vantagens do ELT com Airbyte

- **Agilidade:** Os dados brutos são rapidamente disponibilizados no destino, permitindo transformações flexíveis e iterativas.
- **Flexibilidade:** As transformações podem ser ajustadas dinamicamente para atender às mudanças nas necessidades do negócio.
- **Custos otimizados:** Reduz a necessidade de realizar transformações antes do carregamento, aproveitando a capacidade computacional do Data Warehouse.
- **Open Source e extensível:** Airbyte é uma ferramenta de código aberto, permitindo personalizações e evitando o aprisionamento por fornecedores (vendor lock-in).

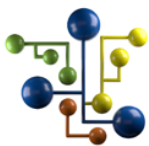
Exemplo Prático

O Airbyte extrai dados de uma API do Shopify e de um banco de dados PostgreSQL.

Esses dados são carregados no Snowflake como tabelas brutas, contendo informações de vendas, clientes e inventário.

Com o dbt, os dados brutos no Snowflake são transformados para criar tabelas modeladas, como relatórios de vendas consolidados e métricas de clientes.

Essa abordagem simplifica o pipeline de dados e maximiza a utilização de recursos modernos, promovendo análises rápidas e eficazes.



Equipe DSA

Muito Obrigado!
Continue Trilhando Uma Excelente Jornada de Aprendizagem.